



Tilburg University

Benchmarking of corporate social responsibility

Graafland, J.J.; Eijffinger, S.C.W.; Smid, H.

Published in:
Journal of Business Ethics

Publication date:
2004

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Graafland, J. J., Eijffinger, S. C. W., & Smid, H. (2004). Benchmarking of corporate social responsibility: Methodological problems and robustness. *Journal of Business Ethics*, 53(1-2), 137-152.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Benchmarking of Corporate Social Responsibility: Methodological Problems and Robustness

Johan J. Graafland
S.C.W. Eijffinger
H. Smid

ABSTRACT. This paper investigates the possibilities and problems of benchmarking Corporate Social Responsibility (CSR). After a methodological analysis of the advantages and problems of benchmarking, we develop a benchmark method that includes economic, social and environmental aspects as well as national and international aspects of CSR. The overall benchmark is based on a weighted average of these aspects. The weights are based on the opinions of companies and NGO's. Using

different methods of weighting, we find that the outcome of the benchmark is rather robust for a sample of more than 50 large Dutch companies.

KEY WORDS: benchmarking, corporate social responsibility, sensitivity analysis, stakeholder perspective, transparency

Johan Graafland studied economics at the Erasmus University Rotterdam. From 1986 he worked at the Central Planning Bureau as an economist. In 1990 he published a dissertation in economics at Erasmus University Rotterdam entitled "Persistent unemployment, wages and hysteresis". From 1993 to 2000 he was head of the division Applied General Equilibrium Models of the Central Planning Bureau. From 1991 to 1998 he studied theology at Utrecht University and wrote a master thesis titled "Economics from a theological perspective: an introduction". In 2000 he was appointed professor in "Economics, business and ethics" at Tilburg University and became Director of the Centre for Corporate Social Responsibility at the faculty of Philosophy. Besides, he worked part-time in the business sector, including family companies in the textile and construction sectors.

Sylvester Eijffinger studied econometrics and economics at the Free University of Amsterdam where he graduated in 1983 and did his PhD in 1986. He was Professor of Economic Policy at the Humboldt University in Berlin from 1995 to 2000 and is still part-time Professor of Monetary Economics at the College of Europe in Bruges. His main position is Professor of Financial Economics and Jean Monnet Professor of European Financial and Monetary Integration at Tilburg University, The Netherlands.

Hugo Smid studies Economics and Philosophy at Tilburg University. In 2002 he became member of the Centre for Corporate Social Responsibility at the Faculty of Philosophy of Tilburg University.

Introduction

The debate in society about the ethical and social aspects of business has forced companies to react to the social and ethical pressure of the public. Many companies nowadays are concerned about values like integrity and feel that they must meet the triple P bottom line expressing the expectations of stakeholders with respect to the company's contribution to profit, planet and people in order to get a licence to operate. Firms that do not meet these expectations may see their reputation founder with a negative impact on market shares and profitability (McIntosh et al., 1998).

The Corporate Social Responsibility (CSR) of a company will only be correctly perceived by the public, if its social and environmental value creation is transparent. One way of improving the transparency of the CSR efforts of companies is benchmarking by independent institutes. Construction of an index that weights the contribution of companies into one number would clarify the position of individual companies and improve the comparability of their CSR efforts. The publication of such an index can potentially enforce the reputation mechanism and provide a competitive advantage to companies that are indeed actively fostering social



and ecological values. This would provide other companies with a strong incentive to integrate CSR into the company's strategy (Graafland, 2002).

Although transparency is an important advantage of benchmarking, ethicists have noticed several methodological problems. These problems relate to the assumption of monism, the assumption of commensurability of various values, the disregard of intentions, the subjectivity of valuation, the notion that the company is more responsible for some of its stakeholders than for others, the assumed context independence of a moral action, the possible lack of control of the company and finally the problem of communication.¹ This paper investigates these problems and develops a method of benchmarking that takes these problems as much as possible into account. For this purpose, we constructed an overall CSR index for more than 50 large Dutch companies based on seventy aspects of CSR by using weights that are derived from the value that companies and NGO's attach to these different aspects of CSR.

The content of this article is as follows. First, the second Section analyzes the advantages of benchmarking. Third Section focuses on the problems involved with benchmarking from a methodological point of view. Fourth Section describes how these methodological weaknesses of benchmarking are dealt with in our approach and compares our approach with other approaches. Fifth Section describes the outcomes and robustness of our benchmark using different methods of weighting. In particular, we compare four benchmarks based on respectively company-based weights, sector-based weights, total averaged-based weights and the weights reported by NGO's, and investigate whether these benchmarks differ significantly. In this way, we also test the relevance of both the sector-specific context of the company as well as the stakeholder-specific context of the benchmark. Sixth section summarizes the main findings.

Advantages of benchmarking CSR

According to the Social Economic Council (2001), CSR incorporates two elements.

- (i) Sufficient focus by the enterprise on its contribution to public prosperity in the longer run.
- (ii) The relationship with its stakeholders and society at large.

The first element stresses that the enterprise can be viewed as a value creating entity. The long term value creation does not only relate to economic value, but concerns value creation in three dimensions which is called the Triple P bottom line

- *Profit*: the economic dimension. This dimension refers to the creation of value through the production of goods and services and through the creation of employment and sources of income. The financial returns reflect the appreciation of consumers for the company's products and the efficiency with which factors of production are used. The long-term profit motive should provide the financial basis for the continuity of the company.
- *People*: the social dimension. This has a variety of aspects concerning the effects for human beings, inside and outside the organization, like good labor relations, safety etc.
- *Planet*: the ecological dimension. This dimension relates to the effects on the natural environment.

Another crucial aspect in this definition is the relationship of the company with its stakeholders and society at large. The enterprise is described as a form of cooperation of different stakeholders. The company should balance between partially conflicting interests. Good stakeholder relations also require that the firm is answering justified questions, opening up its way of doing business, and is willing to have a continuous dialogue with several interested parties. The Social Economic Council distinguishes between primary and other stakeholders. Primary stakeholders are employees and shareholders, who have forms of structural consultation with the managers of the company. Other stakeholders include consumers, suppliers, competitors, the government and society at large, who have a certain interest in the company's activities.

From this definition, it is immediately clear that CSR relates to a set of highly diverse aspects of the behavior of companies. Therefore, it is very difficult to judge a company's performance with respect to CSR. Benchmarking CSR can therefore serve sev-

eral purposes. First, as stressed in the introduction, it can enhance the *transparency*. Through benchmarking, companies are given a mark for their actions and achievements, which enables stakeholders to judge how responsible a specific company is. This will stimulate the stakeholder participation and contribute to an open dialogue. Moreover, while it could benefit the stakeholders, it can also benefit the company itself. It enables the company to signal its CSR efforts. The score is a means for the company to show its stakeholders that it behaves in a responsible way. Without such a benchmark, it is much more difficult. The company itself could report about its responsible behavior, but a benchmark of independent outsiders (academics or agencies) will be much more credible for the stakeholders.

Second, benchmarking improves the *accountability* of the company vis-à-vis its stakeholders. When a score is constructed, it is much easier for stakeholders to confront the company with its actions. The stakeholders can judge whether the company improved its responsibility by comparing this year's score with last year's. The benefit of accountability also applies to the company itself. Through the score, the company is able to identify weaknesses. It could also hold some employees responsible for the CSR achievement. However, an incentive scheme based on a CSR benchmark score could be a questionable practice, because a CSR reporting system may not be as accurate as a financial reporting system.

Third, benchmarking enhances the possibility of cross-company *comparison* with respect to their efforts companies undertake in the light of CSR. A company with a high score is more responsible than a company with a lower score. This enables the stakeholders to compare the various companies and to choose with which company they will involve themselves. The enhanced comparability can also be an advantage for the company itself. On the one hand, it improves the internal comparison, as the company is able to judge the progress compared with last year and enables to identify bottlenecks with respect to CSR. Second, the company can compare its score with the score of other companies. This gives the company a better opportunity to compete with others on the basis of CSR. Benchmarking could eventually result in competition between companies in CSR. Comparability is also important for scientific purposes. For example, it

allows an empirical analysis of the relationship between CSR efforts and other economic factors (profit, etc.), as can be found in Waddock and Graves (1997).

A fourth benefit of the benchmarking procedure is its *simplicity*. Everybody can judge the CSR performance of the company by looking at just one or more scores. In the past, a stakeholder (or company) had to gather a lot of information and make a time-consuming analysis to judge the achievements of the company. Moreover, some data will not be accessible for people who do not belong to the company or a research group. This simplicity implies, however, a rather high level of abstraction. The company will be judged more roughly than when, for example, a story is told about each company, which would enable the reader to form his own opinion. One can say that the researchers who construct the benchmark make choices for the stakeholders who want to judge the company. Stakeholders should thus be aware that these outcomes depend on many assumptions. Therefore, it is very important to explain the exact method used to calculate the benchmark number.

A fifth advantage of benchmarking is that it provides a *systematic approach* to judge the contribution of the company and does not rely on incidents. CSR relates to a set of highly diverse aspects of the behavior of companies. Although companies that spend much attention to CSR have a lower probability on accidents that attract a lot of attention in the media, such events can yield a very unbalanced view on the total achievement of the company. A systematic benchmark of all relevant aspects of the behavior of the company will consequently produce a much more balanced view on the quality of its CSR policy.

Sixth, benchmarking by external independent parties guarantees a more *objective view* than presentation of CSR policies by the company itself. Independence means that the researcher is to pursue the quest for knowledge in a disinterested way. Never is (s)he allowed to let personal interests interfere with the quest for truth as such. Nothing is to be excluded from close scrutiny (De Laat, 2001).

Finally, benchmarking requires that companies are able to deliver the *information* that is needed to construct the benchmark. Companies therefore have an incentive to institutionalize the information

supply process by organizing a systematic database. Once the infrastructure for providing information on CSR efforts is present, it is also easier to realize systematic progress in this field.

Philosophical problems of benchmarking CSR

Notwithstanding the advantages of benchmarking, the method is often criticized by ethicists as well as by entrepreneurs or managers. They point at several problems that especially hold for quantitative versions of the benchmark method. These problems relate to the assumption of monism, the assumption of commensurability of various values, the disregard of intentions, the subjectivity of valuation, the notion that the company is more responsible for some of its stakeholders than for others, the assumed context independence of a moral action, the possible lack of control of the company and finally the problem of communication. We will see that some of these problems are very serious. We need to be aware of these problems in order to realize the relative value of the approach. Also, this analysis will help us to define our own benchmarking approach in such a way as to mitigate the proclaimed disadvantages as much as possible.

Monism and commensurability of values

A benchmark method that expresses the quality of the CSR policy of companies by one single number is monistic in nature. It assumes that it is possible to give a *cardinal* ranking to the realization of different values by different actions. Values are, however, pluralistic in nature. Hurka (1996) defines pluralism as: "A pluralistic theory contains several generic goods rather than only one single one, for example, pleasure, knowledge, and achievement rather than only pleasure.

A monistic theory contains just one generic good". Monism implies that every action can be measured on one single scale, because there is just one good. Hence, we can compare various actions and determine which action generates most value. Moreover, 'good' and 'bad' are symmetrical: one unit of 'good' can offset one unit of 'bad' and only quantity counts, not quality (Van Peperstraten, 1999).

In reality, there is no unique fundamental value for all people. It is much more reasonable to assume pluralism. Van Erp (2000), for instance, mentions: "One can only speak of moral, when different sorts of value are distinguished, some of which are seen as higher than others, or if a priority is given to some values". According to Van Erp, there cannot be monism of values on the empirical level. This is especially true for Kantian ethics. As MacLean (1994) discusses: "Kant famously proclaimed that rational human beings have dignity and that whatever has dignity is 'above all price, and therefore admits of no equivalent'." This implies that the safety of human lives cannot be traded against, for example, profitability.

A somewhat less stringent assumption is to assume that the realization of different values can be ordinally ranked. An *ordinal* ranking says that either value A is better than value B, value B is better than value A, or they are equal in value. Unlike a cardinal ranking, it does not say how much more valuable a value is than the other (Hausman, 1992). Nevertheless, the weaker assumption of ordinal ordering of the overall value of two actions is also problematic for several reasons. First, in the case of an ordinal ranking one has still to assume the commensurability of values. Commensurability means that all values are comparable. Anderson states: "More technically, two goods, A and B, are commensurable if and only if there is a scale of overall value by which they can be at least ordinally ranked." (Anderson, 1993). Such an ordinal ranking is not always possible. Moreover, an ordinal ranking between different values is insufficient for the comparison between the quality of the CSR policy of different companies if the overall quality depends on more than two values. Assume that in an ordinal ranking economic sustainability is considered to be more important than social sustainability and social sustainability more important than ecological sustainability. Suppose further that company A contributes more to economic sustainability, whereas company B performs better in terms of social and ecological sustainability. Then we cannot make an unambiguous judgment about the overall contribution to sustainability of company A as compared to company B, because we do not know whether economic sustainability can also be considered to be more important than social and ecological sustainability together. This problem

becomes even more acute if 50 or 100 different values have to be compared. In order to arrive at an overall judgment, we have to make cardinal rankings that can be used to trade off the performance of A and B for all these values.

Intentions

The benchmarking practice does not deal with intentions. The importance of intentions for a moral evaluation of actions is maybe most clearly stated by Immanuel Kant. Kant's ethical theory is a deontological one. It focuses on the intention of an action (Chryssides and Kaler, 1993). If the intention of a moral action is good, the action itself is morally good. The outcome of that action does not matter.

A good intention also implies that a company should be intrinsically motivated to foster social and ecological values as an end in itself. Intrinsically motivated actions are actions undertaken because the company thinks it is morally obliged to undertake these actions. For instance, it improves the safety of its employees because it thinks it has the moral duty to do that. If attention to the safety of employees is only motivated by the profit motive, a company is only extrinsically interested in its employees as a means for gaining a higher profit and not as an end in itself.

Subjectivity of valuation

In the benchmarking practice, values should be assigned to the actions a company does or does not undertake. For assessing a score, the researcher must get a valuation of each action. How can the researcher obtain such a value for each action? Some stakeholders will value them positively, others negatively. Often cost-benefit analysis is applied to value aspects of policies for which a market is not available. It imitates the market by measuring the values of goods by individuals' "willingness to pay" as a monetary metric for utility. The prices are usually derived from studies of market transactions in which individuals trade off commodified versions of these goods against money. For example, the supposedly higher wage people accept for working at hazardous jobs is used to measure the cash value people are

thought to implicitly place on their own lives. However, it should be noted that very often this type of implicit prices show a very large range because people often make these implicit valuations unconsciously of the implications which economists derive from them. This reduces the usefulness of implicit prices as an objective basis for these calculations. Another problem is that the information required for performing cost-benefit analysis on all the effects of the company is simply not available.

As it is very difficult (and maybe impossible) to be not subjective in the realization of a benchmark score, the score should be presented to the public with care. People should be warned for the fact that there are many subjective elements in the score and that the researchers have already made a lot of choices for them.

Care ethical point of view

Should the company treat all its stakeholders on an equal basis, or should the company take more care for some stakeholders than for others? Whereas the ethics of justice generally takes the former point of view, the ethics of care requires that actions directed at stakeholders to whom the company has a durable and affective relation should be valued more than those directed at other stakeholders (Velasquez, 1998). This implies, for example, that a company has more responsibility to prevent disability of its own employees than to contribute to the reintegration of disabled workers from other companies. In other cases the requirements set by the ethics of care will be company-specific and it will probably be difficult to establish general criteria for determining whether a specific stakeholder needs more care than another.

Context dependence of moral action

The goodness of an action also depends on the context of that action. When a company, for example, faces a decision whether or not to dismiss some of its employees, the moral quality of the decision could not be determined without knowing the decision-making context. For example, for determining the moral value of price agreements between construction companies it is relevant to

know that, if a particular company did not collude, it would have been punished by the other companies and would have had to dismiss people or, worse, would have gone bankrupt. Hence, there are circumstances that diminish the moral blame. When the collusion is placed in this context, it should be morally interpreted in another way than collusion between, for instance, two companies that are able to stop colluding without the chance of going out of business. So, when benchmarking CSR, it is important to judge an action in the context in which it was undertaken. The problem is that it is practically impossible for the benchmarking practice to judge every action within the context in which the action has taken place.

A related complication is that values and norms are often dependent on the company or the sector in which it operates. Take for example a construction company and a retail company. It is plausible that the construction company should take more actions to guarantee healthy working conditions for its employees than a retail company. Therefore, it is problematic to judge a company by looking, for example, at its actions taken to promote work safety, because this norm is sector-specific. This suggests that one should restrict oneself to one sector and cannot apply cross-sector comparisons. Then the assumption of fixed and known values is more plausible.

Control of the company

In the benchmarking method, the company is held responsible for all the actions it does or does not take. However, sometimes the company cannot fully control the effects of its actions. From a moral perspective, a company cannot be held fully responsible for the actions it does not control. As Jeurissen (2000) mentions: “‘must’ in moral sense implies always ‘can’, in practical sense”. The benchmark method does not account for this.

Communication damages moral action

The last problem is that transparency does not always have to be appropriate from a moral perspective. This problem does not only hold for the benchmarking practice as such but for all possible methods

to judge whether a company is responsible or not. In particular, from a Kantian point of view, when one proudly displays CSR actions, this can be seen as damaging the moral status of the action, for it should be not a special thing to do undertake these actions. Therefore, some companies would reject to communicate their moral actions. They would only communicate at times when it is their duty to account for their actions.

This problem has relevance for our research as well. As will be shown in the fifth section, a substantial part of the companies did not want to communicate their ethical standard and also refused to cooperate with this study. How should we interpret this finding? On the one hand, this might indicate a relatively low ethical standard. However, it could also be a deliberate choice of the company not to communicate about its CSR activities, whereas, on the other hand, other companies are perhaps eager to show their CSR activities mainly as a public relations instrument to improve their reputation.

Towards a method of benchmarking

The arguments in third section make clear that benchmarking involves many methodological problems. Does that imply that any benchmark method is useless from a philosophical point of view? In our opinion, this conclusion is too strong. It would imply that one should not make any statement about the ethical quality of one company in relation to that of another company. Although we acknowledge that it is impossible to base the benchmark on an absolutely closed and in every respect satisfactory argumentation, we feel that the problems described in the third section do not prevent us from making deliberate and well-reasoned decisions. In this section we first review some recent benchmark approaches of other researchers. Next, we explain our own benchmark method.

Benchmarking CSR: some recent approaches

Table I gives an overview of three recent approaches of benchmarking CSR.

Krut and Munis (1998) define benchmarking as follows: “A benchmark sets a qualitative stan-

TABLE I
Some recent benchmark methods

	Focus	Quantitative/qualitative	Sources of information	Method of weighting
Krut and Munis (1998)	Environmental performance, 19 categories	Qualitative (typographical symbols)	Statements in public information	No weighting
Graves et al. (2002)	Economic, social and ecological performance, 7 categories	Quantitative (measured on a scale from -2 to +2)	Corporate data sources, questionnaires, periodicals, external surveys	Equal weights
Vlek et al. (2002)	Economic, social and ecological performance	Mixture of added value method and qualitative judgements	Interviews or questionnaires?	Different weights
Our approach	Economic, social and ecological performance, 70 categories	Quantitative (measured on a scale from 0, $\frac{1}{2}$, 1)	Questionnaire plus check by using public information	Weights based on responses of companies and NGO's

dard that allows us to see how a particular firm or a group of firms is moving towards sustainable development". Their method has the following characteristics. First, Krut and Munis focus on 19 categories with respect to the environmental performance of companies, which is more specific than CSR. Krut and Munis concentrate on just one or two sectors, because they acknowledge that comparison of companies within a sector is more valid than comparison of companies between sectors.

Second, their benchmark sets a *qualitative* standard. In particular, Krut and Munis use typographic symbols to show the degree of commitment. In their view, reporting CSR achievements with numbers is inappropriate. By using a qualitative measure, Krut and Munis avoid the problems involved with the quantitative measurement of CSR. However, this comes at the expense of deriving clear conclusions from their standard. Indeed, there is always a trade-off between providing sufficient depth and context for users on the one hand and providing easy comparable outcomes.

Third, Krut and Munis only examine what various companies state about their policies in public information. They do not look at the activities the companies actually undertake. Although they acknowledge that public material could be inflated, they think that this will not be realistic, for com-

panies can be held to their commitments. This assumes sufficient transparency.

Finally, as Krut and Munis only present qualitative judgments, they do not explicitly sum up and weight the scores on the various categories into one overall index.

With the header '100 Best Corporate Citizens' and sub header "America's most responsible and profitable major public companies", the journal Business Ethics opened an article showing the results of its research of the corporate responsibility of various companies that were represented in the S&P 500 (Graves et al., 2002). Graves et al. used various indicators reflecting the performance for the three P's. Stockholder data, for example, are represented by three-year averages of total return to shareholders (capital gains plus dividends). Ecological and social data come from KLD Research and Analytics (Kinder, Lydenberg and Domini, an American statistics bureau) and consist of six categories including pollution emissions, union relationships, employee benefits and philanthropic contributions.

Graves et al. are using a quantitative method. For social data, KLD applies the following method to sum them up. In the six social categories, each company has 'strengths' and 'concerns'. These are rated on scales ranging from -2 (major concern) to 0 (neutral) to +2 (major strength). To arrive at a net score in a category, the concerns are subtracted from

strengths. Thus, a firm with a score of 'two' in employee strengths and 'one' in employee concerns would have a net score of 'one' in the employee category. Notice that in this methodology, 'good' and 'bad' are treated symmetrically.

KLD uses a variety of sources to capture social performance data about each company, like corporate data sources, an annual questionnaire about CSR practices, external data sources, periodicals and external surveys and ratings (Waddock and Graves, 1997). Since the stockholder data and the six social data have different scales, Graves et al. are standardizing these data.

In the final step, they take a non-weighted average of the seven measures to yield a single score for each firm. The fact that the average is non-weighted implies that all stakeholders have equal status. The equal weighting is also an arbitrary choice.

The most comprehensive research we mention is the study by Vlek et al. (2002). Just as Graves et al. (2002), the benchmark method of Vlek et al. focuses on the three Ps: economical value, social value and environmental value. Each of the three dimensions is then described in terms of principles, practices and results of a company. A well-ordered company, as the authors state, is a company where the principles generate practices that should lead to the results the company wants to achieve. Each issue within each of the three dimensions corresponds to principles, and each principle corresponds to a practice and a result. Furthermore, Vlek et al. also distinguish different levels of environmental burden. The lowest level is the direct surroundings of the company. The highest level is the world environmental level. In this way, they take account of the aspect of the control of the company: the influence of one single company diminishes if the level is more global.

In their research, Vlek et al. are using the concept of added value. Like Krut and Munis, they argue that some sorts of CSR achievements will have to be reported qualitatively instead of quantitatively: "A couple of variables, especially social-psychological ones, should not presently be calculated but described". It seems that the reason is a practical one: it would be impossible to measure every aspect quantitatively.

The researchers are not very clear as to how they measure their variables practically. They mention that it can be done with questionnaires or interviews.

The question arises whether it is possible to measure all the variables that are needed in the study by Vlek et al. They have such a comprehensive approach, that it is almost impossible to measure all the relevant variables accurately. Furthermore, because their research is very comprehensive, the benchmarking method itself cannot be very transparent.

Finally, the authors acknowledge that different stakeholders often have different interests: "Stakeholders could judge the importance of specific achievement variables differently". They also acknowledge that the interests of different stakeholders should have different weights in the overall judgment: "The care for the well-being of stakeholders should increase with their dependency and therefore their vulnerability".

Our own approach

In order to counter the methodological problems as well as possible, we develop a benchmark method with the following characteristics.²

- We focus on the three P's and on four sectors (construction, retail, chemical sector and financial and banking sector). We distinguish between 70 aspects or categories of CSR, which are related to six different groups of stakeholders, namely employees, suppliers, customers, society at large, shareholders and competitors.
- For each category, three scores are distinguished (valued respectively by 0, 1/2 and 1.0).
- In order to obtain data of the 70 aspects of CSR, we use a questionnaire. The answers filled in by the companies are checked using annual reports and other public information.
- The most innovative element of our research concerns the second part of the questionnaire, in which the company is asked to give its opinion about the relevance of the different aspects of CSR (they may choose a value of 0, 1/2 or 1.0 for each aspect). Also some Non Governmental Organizations (NGOs) have been sent this part of the questionnaire, which enables us to know how these organizations value the various activities. The respondents

are also asked to attach weights to the relative importance of different stakeholder groups. The overall benchmark is based on a weighted sum of the scores, using the weights given by both the companies and the NGO's. A robustness test and sensitivity analysis is done to examine whether the outcome will be different for various other valuations.

Table II gives an overview of how our approach tries to tackle the problems described in Section 3.

First, although we cannot solve the problem of pluralism in a completely satisfactory way, we use a pragmatic method by defining concrete CSR aspects that entrepreneurs find reasonable and relevant to what they care about. That is one of the reasons why we investigate how much value the respondents and

TABLE II
Methodological problems and mitigating actions in our approach

Problem	Description	Mitigating action in our research
Monism and commensurability	To sum the various values, either monism or commensurability of values should be assumed. These assumptions are, however, doubtful. To calculate a quantitative score, it should be assumed that values are quantifiable	Pragmatic component-value strategy in which the weights are based on the opinion of companies and NGO's
Intentions	The benchmark practice does not deal with intentions, which are important for evaluating a moral action	Assume good intentions are correlated to efforts and procedures
Subjectivity of valuation	Moral actions should be valued, which can for the larger part only be done by the researchers themselves	Selection of aspects and critical values using outcomes of earlier research Ask respondents for weights Perform a robustness test and sensitivity analyses Explain the used method in detail
Care ethical point of view	The practice does not consider that some stakeholders should be given more care than others	Ask respondents for weights per group of stakeholders
Context dependence of moral action	The practice does not consider that valuation of many actions is context dependent and that some norms and values are not fixed	Restrict the research to one sector. Apply the law of large numbers by including many aspects in the benchmark
Control of the company	The practice does not consider that a company that has not full control of some actions cannot be held fully responsible for those actions	Restrict to sectoral comparisons Include efforts (besides outcomes)
Communication damages moral action	The practice needs transparency of the company. The company, however, may not be willing to give information about their moral actions, because this damages the moral status of the action	Only benchmark companies that have filled in the questionnaire

NGO's attach to these different aspects and why we base our weights in the total benchmark on their opinions. As Anderson (1993) argues, such a component-value strategy, in which the overall value is a weighted sum of its component values, can offer a successful solution to the multi-criterion evaluation problem if criteria are relevant for the practice for which they are to be evaluated. It represents evaluation essentially as a matter of calculation, with the aim of making the process precise and decisive. Similar component-value strategies are applied, for example, in decathlon scoring, in which times and distances in different events, such as the hundred-meter dash, the shot put and the long jump, are converted to a common point scale (Anderson, 1993). Basing the weights on the opinions of the companies and NGO's is also consistent with principle P6.3 of the AA1000 standard stating that the identification of indicators that reflect the company's performance must be based on the values of the company itself, the opinions of stakeholders and the society at large (Jonker, 2000).

The second criticism on the lack of a test on moral intentions is encountered by assuming that good intentions will be correlated to CSR efforts. For this purpose, we included several process parameters that measure concrete efforts and procedures to foster ethical standards besides the consequences of these actions in terms of the company's environmental and social performance. This is in accordance with the principle P6.4 of the AA1000 standard (Jonker, 2000). It should be acknowledged that this solution is not fully satisfactory, because efforts and procedures are not necessarily motivated by good will but can also stem from the profit motive.

The subjectivity of valuation is addressed in four ways. First, the initial selection of the aspects to be monitored and the cut-off values used for determining the mark per aspect are based on the outcomes of Graafland et al. (2002), which studies corporate social responsibility for about 100 Dutch companies. This research also focused on the construction, retail and financial and banking sector. Therefore, the outcomes of this research yield good indications about relevant cut-off values determining the valuations per concrete aspect. The subjectivity is further diminished by asking the respondents about how they value the relevance of the concrete

aspects distinguished in the questionnaire. The weights used in the benchmark are based on the average outcomes of this part of the questionnaire. Instead of our own subjective estimates, the benchmark is therefore inter-subjective (although not fully objective). Third, the inter-subjective basis of our questionnaire is tested on its robustness by performing a sensitivity analysis, using four alternative sets of weights: one based on the individual responses of companies, one based on sectoral averages, one based on total averages and one based on the weights of the NGO's. Finally, the objectivity of our benchmark is further raised by keeping it relatively simple and by explaining the method in full detail (Graafland et al., 2003b). This is an advantage of our approach over other more complicated benchmark methods.

Distinguishing explicitly between six different stakeholder groups and asking the respondents about the relative importance of these stakeholder groups deal with the fourth point of criticism. Insofar as care considerations affect the weight attached to certain stakeholder groups, these considerations impact the total outcome of the benchmark.

The fifth point of criticism is dealt with by focusing on specific sectors. Assuming that companies within one sector are subject to similar circumstances, this will enable us at least to compare the benchmark for companies per sector. In order to facilitate comparison of companies belonging to different sectors, we used sector-specific weights reported by the respondents, assuming that these weights correct for different contexts of these companies. Furthermore, by including many aspects of CSR, we hope that the bias caused by very specific contexts for certain aspects of particular companies will not dominate the overall benchmark because of the law of large numbers.

Assuming that companies belonging to a particular sector also face similar problems of lack of control, the sectoral focus of our benchmark approach will also diminish the problem how to correct for lack of control by the company. In addition, this problem is countered by adding categories that measure efforts and procedures rather than the outcome of these efforts. When a company has the intention to eliminate child labor in its company (including its suppliers) and really does much to

eliminate it, but cannot eliminate all child labor because it does not have full control, then the company should not be seen as a less responsible company.

The communication problem is handled by only benchmarking the companies that were prepared to send in the questionnaire. Only if we were to compare companies that have sent in the questionnaire with companies that did not respond, would we somehow have to value the response as such.

The outcomes of the questionnaire

In this section we present the benchmark outcomes of our research. First, we give a short overview of the categories of CSR included in the benchmark and the response to the questionnaire. Next, we present the outcomes of the benchmark and test the robustness using different methods of weighting.

Categories of CSR

Table III presents a condensed overview of the categories used in our benchmark method.

The benchmark comprises CSR aspects relating to economic sustainability (in particular, prevention of inside trading, profitability, measures to prevent collusion and bribery), ecological sustainability (like ecological effects of the production process and product of the company itself as well as of its suppliers) and social sustainability (such as the labor conditions of the company itself and its suppliers, the contribution to social problems of society at large). In addition, like Kleinfeld (2001), we include the use of several instruments that facilitate responsible behavior of the company, like the ISO standards, code of conduct, ethical committee and ethical training (Graafland et al. 2003a).

Response

The addresses of the companies were taken from publications of PriceWaterhouseCoopers (2001, 2002a–c). These publications also contain detailed statistics about the number of employees and the profitability of the company. In total we sent our

questionnaire to 378 companies. The response rate varied from 20% in the construction and chemical sector to 14% for the financial and banking sector and 9% for the retail sector. On average, the response rate was 15.3%.

In order to check the reliability of the answers, we checked the data by researching annual reports, newspaper articles and Internet sites to detect relevant information for a subset of 30 aspects. On average, we found divergences for only 1 or 2 aspects per company. Since there was no systematic upward or downward bias in the answers of the companies, we decided not to adapt the answers filled in by the companies themselves.

Four methods of weighting

The overall benchmark (B) was constructed by using the following formula:

$$B = (\text{Sum}(w_i \times b_i) + w_o b_o) / (\text{Sum}(w_i) + w_o)$$

w_i denotes the weight per stakeholder and w_o is the weight for the use of instruments. b_i is the benchmark per stakeholder and b_o the benchmark for the use of instruments. b_i is constructed by the following formula:

$$b_i = \text{Sum}(w_j \times b_j) / \text{Sum}(w_j)$$

where w_j denotes the weight per aspect of CSR and b_j the value of the option per category filled in by the company. The benchmark for the use of instruments is calculated in a similar way.

As we asked the opinion of the respondents and NGO's about the weights to be used for the benchmark, we have four alternatives of weighting. First, we can construct an overall benchmark index using for each company the weights reported by the companies themselves. The company then defines by itself the criteria that are relevant to judge its CSR performance. We will label the benchmarks with this weighting scheme the *individual* benchmarks. The advantage of this method of weighting is that the benchmark will maximally reflect the context that the individual company perceives as most relevant. The disadvantage of this method is that the weight might be biased. In particular, companies might attach high values to CSR aspects for which they rank high.

TABLE III
Categories and aspects of corporate social responsibility per stakeholder

<i>Employees</i>	Human rights	Dialogue with NGO's	Equal opportunity for women	Equal opportunity for minorities	
	Training	Safety and health	Participation	Correct attitude	Good fellowship among workers
<i>Suppliers</i>	Safety of product	Environmental effect of product and production process	Labor conditions of supplier	Respect for supplier	
<i>Customers</i>	Safety and quality of product	Supply of sustainable alternative	Respect for customer		
<i>Society at large</i>	Environmental effects	Active dialogue with environmental organizations	Reintegration of disabled people	Contribution to reduction of poverty in third world	Contribution to local projects
<i>Shareholders</i>	Prevention of inside trade of stocks	Profitability			
<i>Competitors</i>	Respect for intellectual property of competitors	Measures to prevent collusion	Measures to prevent bribery		
<i>Organization of ethics</i>	Various instruments, including code of conduct, ISO certification, external audits, social reports, social handbook, ethical committee and ethical training				

A more objective method is the second approach in which we used averaged weights per *sector*. As companies operating in one particular sector face similar conditions, using sectoral weights will still take account of the context of the company, although not as specific as in the case of individual weights. We will name these benchmarks the *sectoral (based)* benchmarks. An advantage of this approach is that subjective judgments and the upward bias created by using individual weights will be filtered out.

A third method is to use average weights based on the *total* sample of companies. We will call these benchmarks the *total average* or *total based* benchmarks. Compared to the second approach, this method of weighting does not take account of context aspects related to typical sectoral characteristics. Still, it is interesting to test whether this general method of weighting produces completely different benchmark results.

In the last approach we will use the average weights of the *NGO's*. These benchmarks will be referred to as the *NGO (based)* benchmarks. This also provides insight into the robustness of the benchmark approach. In particular, it allows us to test how stakeholder related the benchmark is. Of course, we are very much interested to see whether and to which degree the individual, sectoral, average and NGO benchmarks will diverge from each other.

Robustness of the benchmarks

Figures 1–3 compare the outcomes of the various benchmarks. These figures show that the differences between the sectoral benchmark, the NGO benchmark and the average benchmark are not large. In contrast, if the total average benchmark and individual benchmark are plotted together the points are less close to the reference line (45° line), which means that the differences between the two benchmarks are larger. The picture also makes clear that the difference between the benchmark based on individual weights and the benchmark based on total average weights is more likely to be positive for companies with a high CSR performance than for companies with a low CSR performance. This might indicate that companies with a high CSR performance are more aware which aspects of CSR receive a high valuation in their company. As a result, they are more able to provide a valuation that is consistent with their performance.

On basis of these results, we conclude that the weights do not generally have a strong impact on the benchmarks. The ascending order is rather robust. In all sectors, the company receiving the highest mark is the same for weights based on sectoral averages, total averages or NGO's averages. This has important implications

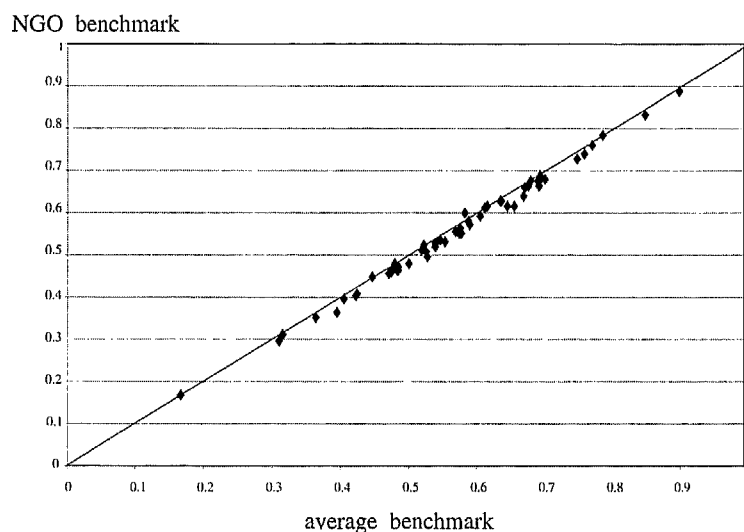


Figure 1. NGO benchmark plotted with the total average benchmark.

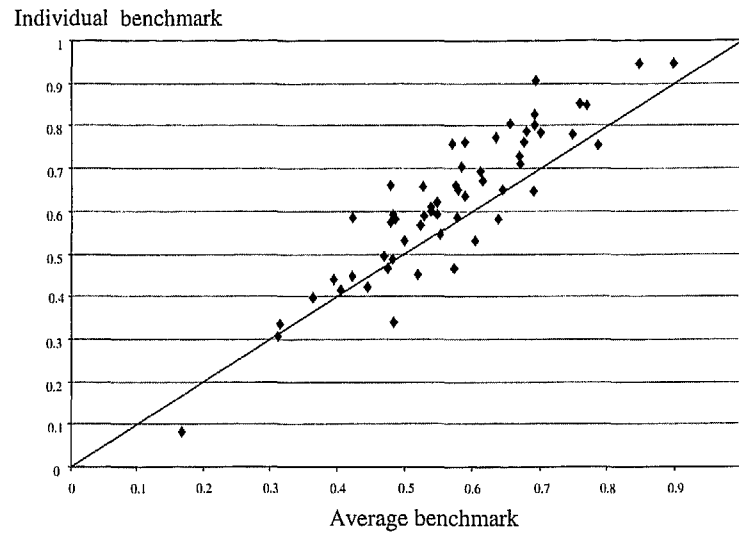


Figure 2. Individual benchmark plotted with the total average benchmark.

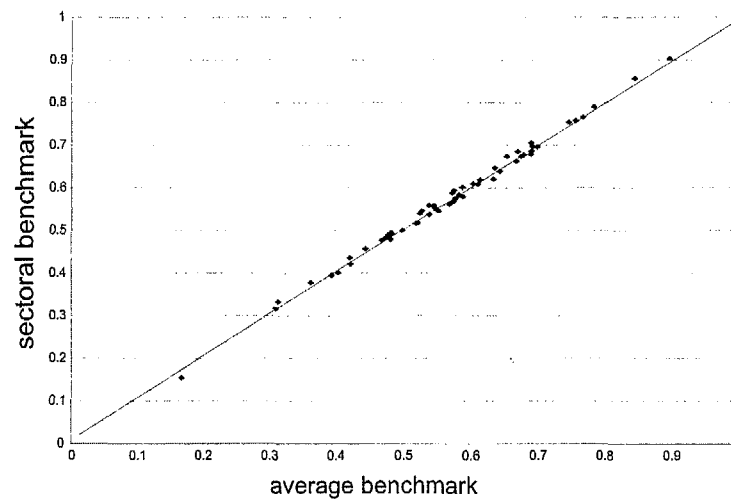


Figure 3. Sectoral benchmark plotted with the total average benchmark.

- (i) the sectoral-specific context is not of decisive importance, since *sectoral* based weights hardly change the benchmark compared to total based weights;
- (ii) also stakeholder-specific opinions hardly matter, as the *NGO* based weights do not impact the benchmark outcome compared to total based weights.

Only in the case of weights based on the *individual* answers of companies, the company receiving the highest mark is not the same as the one selected by

other methods. Since individual weights are relatively subjective and might be biased to the actual performance of companies, the weights based on sectoral averages may provide a more reliable guide since it still takes into account the sectoral context whilst being more inter-subjective in nature.

Summary of the main findings

There are many advantages to benchmarking of corporate social responsibility of companies. For

example, benchmarking serves transparency. Through benchmarking, companies are given a mark for their actions and achievements, which enables stakeholders to judge how responsible a specific company is. When a score is constructed, it is much easier for stakeholders to confront the company with its actions. This increases its accountability in terms of CSR.

Notwithstanding these and other advantages, benchmarking of CSR is also highly criticized because it tends to disregard the complexity of measuring responsible behavior. These problems relate to the assumption that values can be reduced to one dimension (monism) and that values are comparable (commensurability). Moreover, benchmarking may disregard the intentions of the company or the particular context in which the company operates. Furthermore, the estimation of the quality of the company's CSR efforts may be rather subjective because of measurement problems, whereas the possible lack of control of the company cannot be taken into account either.

In order to encounter the problems as well as possible we have developed a benchmark method that focuses on economic, social and environmental value creation in four Dutch sectors (construction, retail, chemical sector and financial and banking sector). We have sent a questionnaire to Dutch companies with 70 aspects of CSR, which are both related to the effort and procedures of the company to foster ethical standards as well as to the results of these efforts. We have checked the answers using annual reports and other sources of public information. The weights of these categories in the overall benchmark are based on the company's opinion about the relevancy of these different aspects of CSR. Also some NGO's have been sent this part of the questionnaire in order to know how these organizations value the various activities.

Comparison of the four benchmarks shows that the various weights do not have a strong impact on the benchmarks. The ascending order is rather robust. In all sectors, the company receiving the highest mark is the same for weights based on the sectoral averages, total averages or NGO's averages. Also the top five in the total ordering of all companies is very robust when these different weights are used. This indicates that neither the sectoral context nor specific stakeholder perceptions have a major

impact on the benchmark results. Only in the case of weights based on the individual answers of companies, the company receiving the highest mark differs from the one selected by the other methods. Since individual weights are relatively subjective and might be biased to the actual performance of companies, the weights based on sectoral averages may provide a more reliable guide because it still takes into account the sectoral context, whilst being more inter-subjective in nature.

Acknowledgement

The authors thank N.C.G.M. Stoffele and A.M. Coldeweijer for their assistance in the research.

Notes

¹ These problems are also apparent in benchmarking, for example, the transparency of central banks regarding monetary policymaking. See e.g. Eijffinger and Geraats (2002).

² For a more detailed report of our benchmark method and outcomes, see Graafland et al. (2003).

References

- Anderson, E.: 1993, *Value in Ethics and Economics* (Harvard University Press, Cambridge/London).
- Chryssides, G. and J. H. Kaler: 1993, *An Introduction to Business Ethics* (Chapman & Hall, London).
- Erp, H. van: 2000, *Moraal geluk en verantwoordelijkheid* (Van Gorcum, Assen).
- Eijffinger, S. C. W. and P. M. Geraats: 2002, 'How Transparent are Central Banks?', CEPR Discussion Paper, Centre for Economic Policy Research, London.
- Graafland, J. J.: 2002, 'Modelling the Trade off between Profits and Principles', *De Economist* **150**, 129–154.
- Graafland, J. J., B. W. van de Ven and N. Stoffele: 2002, 'Wat betekent maatschappelijk ondernemen concreet?', University of Tilburg, www.uvt.nl/wijsbegeerte/cmo
- Graafland, J. J., B. W. van de Ven and N. Stoffele: 2003a, 'Strategies and instruments for organising CSR by small and large business in the Netherlands', *Journal of Business Ethics* **47**, 45–60.

- Graafland, J. J., S. C. W. Eijffinger, N. C. G. M. Stoffele, H. Smid and A. M. Coldeweijer: 2003b, 'Corporate social responsibility of Dutch companies: Benchmarking and Transparency', <http://www.uvt.nl/faculteiten/fww/cmo>.
- Graves, S. B., S. A. Waddock and J. Kelly: 2002, '100 Best Corporate Citizens', *Business Ethics* **11**(2), 8–13.
- Hausman, D. M.: 1992, *The Inexact and Separate Science of Economics* (Cambridge University Press, Cambridge).
- Hurka, T.: 1996, 'Monism, Pluralism, and Rational Regret', *Ethics: An International Journal of Social, Political and Legal Philosophy* **106**(3), 555–575.
- Jeurissen, R. J. M. (ed.): 2000, *Bedrijfsethiek: een goed zaak* (Van Gorcum, Assen).
- Jonker, J. (ed.): 2000, *AA 1000 standaard: richtlijn voor maatschappelijk verantwoord ondernemen* (Van Gorcum, Assen).
- Kleinfeld, A.: 2001, 'Benchmarking the moral decision-making strength of European biotech companies: a European research project', *Business Ethics: A European Review* **10**, 122–139.
- Krut, R. and K. Munis: 1998, 'Sustainable Industrial Development: Benchmarking Environmental Policies and Reports', *Greener Management International* **21**, 88–98.
- Laat, P. B. de: 2001, 'Open source software: A new Mertonian ethos?', in A. Vedder (ed.), *Ethics and the Internet* (Intersentia, Antwerpen-Groningen-Oxford), pp. 33–47.
- MacLean, D.: 1994, 'Cost-Benefit Analysis and Procedural Values', *Analyse & Kritik*, **16**(2), 166–180.
- McIntosh, M., D. Leipziger, K. Jones and G. Coleman: 1998, *Corporate Citizenship: Successful Strategies for Responsible Companies* (Financial Times Pitman Publishing, London/San Francisco).
- Peperstraten, F. van: 1999, *Samenleving ter discussie* (Coutinho, Bussum).
- PriceWaterhouseCoopers: Ondernemingsanalyses: 2001, *Trends in de chemie* (Elsevier bedrijfsinformatie B. V., Dordrecht).
- PriceWaterhouseCoopers: Ondernemingsanalyses: 2002a, *Trends in de bouw* (Elsevier bedrijfsinformatie B. V., Dordrecht).
- PriceWaterhouseCoopers: Ondernemingsanalyses: 2002b, *Trends in de retail* (Elsevier bedrijfsinformatie B. V., Dordrecht).
- PriceWaterhouseCoopers: Ondernemingsanalyses: 2002c, *Trends bij banken* (Elsevier bedrijfsinformatie B. V., Dordrecht).
- Social Economic Council: 2001, *Corporate social Responsibility: A Dutch Approach* (Van Gorcum, Assen).
- Velasquez, M. G.: 1998, *Business ethics. Concepts and Cases*, 4th Edition (Prentice Hall, Upper Saddle River, New Jersey).
- Vlek, C. A. J., E. M. Steg, D. Feenstra, W. Gerbens-Leenis, S. Lindenberg, H. Moll, A. Schoot Uiterkamp, F. Sijtsma and A. Van Witteloostuijn: 2002, 'Een praktisch model voor duurzaam bedrijfspresteren', *Economisch Statistische Berichten* **87**, 524–527.
- Waddock, S. A. and S. B. Graves: 1997, 'The corporate social performance-financial performance link', *Strategic Management Journal*, 303–319.

Tilburg University,
Warandelaan 2,
P.O. Box 90153,
5000 LE Tilburg,
The Netherlands,
E-mail: J.J.Graafland@uvt.nl